

862.C2178



PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

YASUO OKUTANI, ET AL.

Application No.: 09/819,613

Filed: March 29, 2001

For: SPEECH SIGNAL PROCESSING APPARATUS  
AND METHOD, AND STORAGE MEDIUM

)  
:  
Examiner: Not Assigned

)  
:  
Group Art Unit: 1762

)  
:  
July 16, 2001

Commissioner for Patents  
Washington, D.C. 20231

RECEIVED  
JUL 18 2001  
TC 1700

CLAIM TO PRIORITY

Sir:

Applicants hereby claim priority under the International Convention and all rights to which they are entitled under 35 U.S.C. § 119 based upon the following Japanese Priority Application:

JAPAN

2000-099532

March 31, 2001.

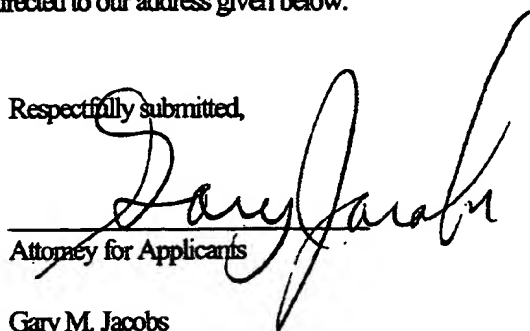
A certified copy, with the translation of the first page of the priority document, is enclosed.

BEST AVAILABLE COPY

Applicants' undersigned attorney may be reached in our Washington, D.C. office by telephone at

(202) 530-1010. All correspondence should continue to be directed to our address given below.

Respectfully submitted,



Attorney for Applicants

Gary M. Jacobs

Registration No. 28,861

FITZPATRICK, CELLA, HARPER & SCINTO

30 Rockefeller Plaza

New York, New York 10112-3801

Facsimile: (212) 218-2200

GMJ/cm

DC\_MAIN 83766 v

BEST AVAILABLE COPY

01/01/01  
Yasuo OKUTANI, et al  
32917

(translation of the front page of the priority document of  
Japanese Patent Application No. 2000-099532)



PATENT OFFICE  
JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the  
following application as filed with this Office.

Date of Application: March 31, 2000

Application Number : Patent Application 2000-099532

Applicant(s) : Canon Kabushiki Kaisha

RECEIVED

JUL 18 2001

TC 1700

April 20, 2001

Commissioner,  
Patent Office

Kouzo OIKAWA

Certification Number 2001-3033159

BEST AVAILABLE COPY

日本国特許庁  
JAPAN PATENT OFFICE

CIFM 2178 US  
09/8/19, 613  
Yasuo OKUTANI, claf  
3-29-01

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日  
Date of Application:

2000年 3月31日

出願番号  
Application Number:

特願2000-099532

出願人  
Applicant(s):

キヤノン株式会社



RECEIVED

JUL 18 2001

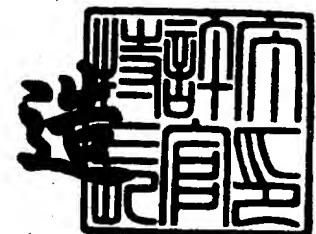
TC 1700

CERTIFIED COPY OF  
PRIORITY DOCUMENT

2001年 4月20日

特許庁長官  
Commissioner,  
Japan Patent Office

及川耕造



BEST AVAILABLE COPY

出証番号 出証特2001-3033159

【書類名】 特許願

【整理番号】 4172017

【提出日】 平成12年 3月31日

【あて先】 特許庁長官殿

【国際特許分類】 G01L 5/04

【発明の名称】 音声情報処理装置及びその方法と記憶媒体

【請求項の数】 23

【発明者】

    【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社  
社内

    【氏名】 奥谷 泰夫

【発明者】

    【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社  
社内

    【氏名】 小森 康弘

【発明者】

    【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社  
社内

    【氏名】 深田 俊明

【特許出願人】

    【識別番号】 000001007

    【氏名又は名称】 キヤノン株式会社

【代理人】

    【識別番号】 100076428

    【弁理士】

    【氏名又は名称】 大塚 康德

    【電話番号】 03-5276-3241

【選任した代理人】

    【識別番号】 100101306

【弁理士】

【氏名又は名称】 丸山 幸雄

【電話番号】 03-5276-3241

【選任した代理人】

【識別番号】 100115071

【弁理士】

【氏名又は名称】 大塚 康弘

【電話番号】 03-5276-3241

【手数料の表示】

【予納台帳番号】 003458

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0001010

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声情報処理装置及びその方法と記憶媒体

【特許請求の範囲】

【請求項 1】 音韻環境を満足する音声素片を音声データベースから検索する音声素片検索手段と、

前記音声素片検索手段による検索結果に基づいて各音韻の HMM モデルを算出する HMM 学習手段と、

前記各音韻の HMM モデルに基づいて前記音声素片の素片認識を行なう素片認識手段と、

前記素片認識手段による素片認識結果に応じて素片辞書に登録する音声素片を決定する登録素片決定手段と、

を有することを特徴とする音声情報処理装置。

【請求項 2】 前記素片認識手段は、前記音韻の単位としてダイフォンを採用し、音声素片を CC 型、CV 型、VC 型、VV 型の 4 種類に分離して（C：子音、V：母音）、前記型のそれぞれにおいて素片認識を行なうことを特徴とする請求項 1 に記載の音声情報処理装置。

【請求項 3】 前記登録素片決定手段は、

許容できる音声素片パターンに登録したパターン記憶手段を有し、

前記素片認識手段において素片認識に失敗した音声素片に一致する音声素片パターンが前記パターン記憶手段に登録されているか否かを判定し、登録されている場合に前記音声素片を前記素片辞書に登録することを特徴とする請求項 1 又は 2 に記載の音声情報処理装置。

【請求項 4】 前記登録素片決定手段は、前記素片認識手段により認識した音声素片の数が所定値以上である場合に前記素片辞書に登録することを特徴とする請求項 1 に記載の音声情報処理装置。

【請求項 5】 前記登録素片決定手段は、前記素片認識手段により認識した音声素片の数が前記所定数以下であっても、少なくとも母音部分が正しく認識されていた場合は前記素片辞書に登録することを特徴とする請求項 4 に記載の音声情報処理装置。

【請求項 6】 前記素片認識手段は、音韻が同じ音声素片を対象に尤度を求め、更に、

前記登録素片決定手段は、前記尤度が上位のもの、或いは、所定値以上の音声素片を前記素片辞書に登録することを特徴とする請求項 1 に記載の音声情報処理装置。

【請求項 7】 前記登録素片決定手段は、前記尤度を前記音声素片の時間長で正規化した値が上位のもの、或いは、所定値以上の音声素片を前記素片辞書に登録することを特徴とする請求項 6 に記載の音声情報処理装置。

【請求項 8】 音韻環境を満足する音声素片を音声データベースから検索する音声素片検索工程と、

前記音声素片検索工程における検索結果に基づいて各音韻の HMM モデルを算出する HMM 学習工程と、

前記各音韻の HMM モデルに基づいて前記音声素片の素片認識を行なう素片認識工程と、

前記素片認識工程による素片認識結果に応じて素片辞書に登録する音声素片を決定する登録素片決定工程と、

を有することを特徴とする音声情報処理方法。

【請求項 9】 前記素片認識工程は、前記音韻の単位としてダイフオンを採用し、音声素片を CC 型、CV 型、VC 型、VV 型の 4 種類に分離して（C：子音、V：母音）、前記型のそれぞれにおいて素片認識を行なうことを特徴とする請求項 8 に記載の音声情報処理方法。

【請求項 10】 前記登録素片決定工程は、

許容できる音声素片パターンに登録したパターン記憶工程を有し、

前記素片認識工程において素片認識に失敗した音声素片に一致する音声素片パターンが前記パターン記憶工程に登録されているか否かを判定し、登録されている場合に前記音声素片を前記素片辞書に登録することを特徴とする請求項 8 又は 9 に記載の音声情報処理方法。

【請求項 11】 前記登録素片決定工程では、前記素片認識工程により認識した音声素片の数が所定値以上である場合に前記素片辞書に登録することを特徴



とする請求項 8 に記載の音声情報処理方法。

【請求項 1 2】 前記登録素片決定工程では、前記素片認識工程により認識した音声素片の数が前記所定数以下であっても、少なくとも母音部分が正しく認識されていた場合は前記素片辞書に登録することを特徴とする請求項 1 1 に記載の音声情報処理方法。

【請求項 1 3】 前記素片認識工程では、音韻が同じ音声素片を対象に尤度を求め、更に、

前記登録素片決定工程では、前記尤度が上位のもの、或いは、所定値以上の音声素片を前記素片辞書に登録することを特徴とする請求項 8 に記載の音声情報処理方法。

【請求項 1 4】 前記登録素片決定工程では、前記尤度を前記音声素片の時間長で正規化した値が上位のもの、或いは、所定値以上の音声素片を前記素片辞書に登録することを特徴とする請求項 1 3 に記載の音声情報処理方法。

【請求項 1 5】 請求項 8 乃至 1 4 のいずれか 1 項に記載の方法を実行するプログラムを記憶したことを特徴とする、コンピュータにより読取り可能な記憶媒体。

【請求項 1 6】 請求項 8 乃至 1 4 のいずれか 1 項に記載の方法により音声素片が登録された素片辞書と、

入力されたテキストデータの言語解析を行なう言語解析手段と、

前記言語解析手段により解析された解析結果に基づいて韻律を生成する韻律生成手段と、

前記韻律生成手段により生成された韻律を基に前記素片辞書を検索して対応する音声素片を選択する音声素片選択手段と、

前記音声素片選択手段により選択された音声素片を編集して接続する音声素片編集・接続手段と、

前記音声素片編集・接続手段により編集された結果に基づいて音声を再生する音声再生手段と、

を有することを特徴とする音声情報処理装置。

【請求項 1 7】 所定の音韻環境を満足する複数の音声素片を用いて、各音

韻に対応するHMMを学習するHMM学習手段と、

前記各音韻に対応するHMMに基づいて、音声合成に使用する素片辞書に登録する音声素片を選択する登録素片決定手段と、  
を有することを特徴とする音声情報処理装置。

【請求項18】 前記登録素片決定手段は、前記複数の音声素片の一音声素片と最も尤度の高くなる最尤HMMを前記各音韻に対応するHMMの中から求め、前記一音声素片が前記最尤HMMの学習に使用した音声素片か否かを判定し、前記最尤HMMの学習に使用した音声素片である場合に前記一音声素片を選択することを特徴とする請求項17に記載の音声情報処理装置。

【請求項19】 前記素片辞書を用いて音声合成する音声合成手段を更に有することを特徴とする請求項17又は18に記載の音声情報処理装置。

【請求項20】 所定の音韻環境を満足する複数の音声素片を用いて、各音韻に対応するHMMを学習するHMM学習工程と、

前記各音韻に対応するHMMに基づいて、音声合成に使用する素片辞書に登録する音声素片を選択する登録素片決定工程と、  
を有することを特徴とする音声情報処理方法。

【請求項21】 前記登録素片決定工程では、前記複数の音声素片の一音声素片と最も尤度の高くなる最尤HMMを前記各音韻に対応するHMMの中から求め、前記一音声素片が前記最尤HMMの学習に使用した音声素片か否かを判定し、前記最尤HMMの学習に使用した音声素片である場合に前記一音声素片を選択することを特徴とする請求項20に記載の音声情報処理方法。

【請求項22】 前記素片辞書を用いて音声合成する音声合成工程を更に有することを特徴とする請求項20又は21に記載の音声情報処理方法。

【請求項23】 請求項20乃至22のいずれか1項に記載の方法を実行するプログラムを記憶したことを特徴とする、コンピュータにより読取り可能な記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声合成で使用される素片辞書を作成する音声情報処理装置及びその方法と記憶媒体に関するものである。

【 0 0 0 2 】

【従来の技術】

近年、音素やd i p h o n eなどを単位とした音声素片を素片辞書に登録しておき、音声合成に際して、入力される表音テキストに従って素片辞書を検索し、それら検索された音声素片を編集、接続して、その表音テキストに対応する音声合成を行なって音声を出力する音声合成方式が主流となっている。

【 0 0 0 3 】

【発明が解決しようとする課題】

このような音声合成方式においては、合成音声の基になる、素片辞書に登録されている音声素片自体の精度が重要となる。従って、音声素片の音韻環境にバラツキがあったり、或いは音声素片が雑音を含んでいると、いくら優れた精度の音声合成を行なっても、それら音声素片を用いて生成される合成音声には異音や雑音が含まれてしまうことになる。

【 0 0 0 4 】

本発明は上記従来例に鑑みてなされたもので、HMMを用いた素片認識を行ない、その認識結果に応じて辞書に登録する音声情報処理装置及びその方法と記憶媒体を提供することを目的とする。

【 0 0 0 5 】

また本発明の目的は、音声合成における音質の劣化を防止した素片辞書を作成する音声情報処理装置及びその方法と記憶媒体を提供することにある。

【 0 0 0 6 】

【課題を解決するための手段】

上記目的を達成するために本発明の音声情報処理装置は以下のような構成を備える。即ち、

音韻環境を満足する音声素片を音声データベースから検索する音声素片検索手段と、

前記音声素片検索手段による検索結果に基づいて各音韻のHMMモデルを算出

する HMM 学習手段と、

前記各音韻の HMM モデルに基づいて前記音声素片の素片認識を行なう素片認識手段と、

前記素片認識手段による素片認識結果に応じて素片辞書に登録する音声素片を決定する登録素片決定手段と、  
を有することを特徴とする。

【 0 0 0 7 】

上記目的を達成するために本発明の音声情報処理方法は以下のような工程を備える。即ち、

音韻環境を満足する音声素片を音声データベースから検索する音声素片検索工程と、

前記音声素片検索工程における検索結果に基づいて各音韻の HMM モデルを算出する HMM 学習工程と、

前記各音韻の HMM モデルに基づいて前記音声素片の素片認識を行なう素片認識工程と、

前記素片認識工程による素片認識結果に応じて素片辞書に登録する音声素片を決定する登録素片決定工程と、  
を有することを特徴とする。

【 0 0 0 8 】

上記目的を達成するために本発明の音声情報処理装置は以下のような構成を備える。即ち、

請求項 8 乃至 1 4 のいずれか 1 項に記載の方法により音声素片が登録された素片辞書と、

入力されたテキストデータの言語解析を行なう言語解析手段と、

前記言語解析手段により解析された解析結果に基づいて韻律を生成する韻律生成手段と、

前記韻律生成手段により生成された韻律を基に前記素片辞書を検索して対応する音声素片を選択する音声素片選択手段と、

前記音声素片選択手段により選択された音声素片を編集して接続する音声素片

編集・接続手段と、

前記音声素片編集・接続手段により編集された結果に基づいて音声を再生する音声再生手段と、

を有することを特徴とする。

また、本発明の音声情報処理装置は以下のような構成を備える。即ち、

所定の音韻環境を満足する複数の音声素片を用いて、各音韻に対応するHMMを学習するHMM学習手段と、

前記各音韻に対応するHMMに基づいて、音声合成に使用する素片辞書に登録する音声素片を選択する登録素片決定手段と、

を有することを特徴とする。

#### 【 0 0 0 9 】

##### 【発明の実施の形態】

以下、添付図面を参照して本発明の好適な実施の形態を詳細に説明する。

#### 【 0 0 1 0 】

##### 〔実施の形態 1〕

図 1 は、本発明の実施の形態に係る音声合成装置のハードウェア構成を示すブロック図である。尚、本実施の形態では、一般的なパーソナルコンピュータを音声合成装置として用いる場合について説明するが、本発明は専用の音声合成装置であっても、また他の形態の装置であっても良い。

#### 【 0 0 1 1 】

図 1 において、101は制御メモリ（ROM）で、中央処理装置（CPU）102で使用される各種制御データを記憶している。CPU102は、RAM103に記憶された制御プログラムを実行して、この装置全体の動作を制御している。103はメモリ（RAM）で、CPU102による各種制御処理の実行時、ワークエリアとして使用されて各種データを一時的に保存するとともに、CPU102による各種処理の実行時、外部記憶装置104から制御プログラムをロードして記憶している。この外部記憶装置は、例えばハードディスク、CD-ROM等を含んでいる。105はD/A変換器で、音声信号を示すデジタルデータが入力されると、これをアナログ信号に変換してスピーカ109に出力して音声を再

生する。106は入力部で、オペレータにより操作される、例えばキーボードや、マウス等のポインティングデバイスを備えている。107は表示部で、例えばCRTや液晶等の表示器を有している。108はバスで、これら各部を接続している。110は音声合成ユニットである。

#### 【0012】

以上の構成において、本実施の形態の音声合成ユニット110を制御するための制御プログラムは外部記憶装置104からロードされてRAM103に記憶され、その制御プログラムで用いる各種データは制御メモリ101に記憶されている。これらのデータは、中央処理装置102の制御の下にバス108を通じて適宜メモリ103に取り込まれ、中央処理装置102による制御処理で使用される。D/A変換器105は、制御プログラムを実行することによって作成される音声波形データ(デジタル信号)をアナログ信号に変換してスピーカ109に出力する。

#### 【0013】

図2は、本実施の形態に係る音声合成ユニット110のモジュール構成を示すブロック図で、この音声合成ユニット110は、大きく分けて、素片辞書206に音声素片を登録するための処理を実行する素片辞書作成モジュールと、テキストデータを入力し、そのテキストデータに対応する音声を合成して出力する処理を行なう音声合成モジュールの2つのモジュールを有している。

#### 【0014】

図2において、201は、入力部106や外部記憶装置104から任意のテキストデータを入力するテキスト入力部、202は解析辞書、203は言語解析部、204は韻律生成規則保持部、205は韻律生成部、206は素片辞書、207は音声素片選択部、208はPSOLA法(ピッチ同期波形重畳法)を用いて音声素片を編集する音声素片編集・接続部、209は音声波形出力部、210は音声データベース、211は素片辞書作成部である。

#### 【0015】

まず音声合成モジュール処理について説明する。音声合成モジュールでは、言語解析部203が、解析辞書202を参照して、テキスト入力部201から入力

されるテキストの言語解析を行なう。こうして解析された結果が韻律生成部 2 0 5 に入力される。韻律生成部 2 0 5 は、言語解析部 2 0 3 における解析結果と、韻律生成規則保持部 2 0 4 に保持されている韻律生成規則に関する情報とを基に音韻系列と韻律情報とを生成して音声素片選択部 2 0 7 及び音声素片編集・接続部 2 0 8 に出力する。続いて、音声素片選択部 2 0 7 は、韻律生成部 2 0 5 から入力される韻律生成結果を用いて、素片辞書 2 0 6 に保持されている音声素片から対応する音声素片を選択する。音声素片編集・接続部 2 0 8 は、韻律生成部 2 0 5 から入力される韻律生成結果に従って、音声素片選択部 2 0 7 から出力される音声素片を編集及び接続して音声波形を生成する。こうして生成された音声波形は、音声波形出力部 2 0 9 で出力される。

## 【 0 0 1 6 】

次に、素片辞書作成モジュールについて説明する。

## 【 0 0 1 7 】

このモジュール処理では、素片辞書作成部 2 1 1 が、後述する手順に基づいて音声データベース 2 1 0 の中から音声素片を選び出して素片辞書 2 0 6 に登録する。

## 【 0 0 1 8 】

次に、上記構成を備えた本実施の形態の音声合成処理について説明する。

## 【 0 0 1 9 】

図 3 は、図 2 の音声合成モジュールにおける音声合成処理(オンライン処理)の流れを示すフローチャートである。

## 【 0 0 2 0 】

まずステップ S 3 0 1 で、テキスト入力部 2 0 1 は、文、文節、単語などの単位毎にテキストデータを入力してステップ S 3 0 2 に移る。ステップ S 3 0 2 では、言語解析部 2 0 3 により当該テキストデータの言語解析を行う。次にステップ S 3 0 3 に進み、音韻生成部 2 0 5 は、ステップ S 3 0 2 で解析された結果と所定の韻律規則とに基づいて、音韻系列と韻律情報を生成する。次にステップ S 3 0 4 に進み、各音韻毎にステップ S 3 0 3 で得られた韻律情報と所定の音韻環境とに基づいて、音声素片選択部 2 0 7 が素片辞書 2 0 6 に登録されている音声

素片を選択する。次にステップS305に進み、その選択された音声素片及びステップS303で生成された韻律情報とに基づいて、音声素片編集・接続部208により音声素片の編集および接続を行なってステップS306に進む。ステップS306では、音声素片編集・接続部208によって生成された音声波形を、音声波形出力部209が音声信号として出力する。このようにして、入力されたテキストに対応する音声出力されることになる。

#### 【0021】

図4は、図2で示した素片辞書作成モジュールのより詳細な構成を示すブロック図で、前述の図2と共通する部分は同じ番号で示し、かつ本実施の形態の特徴である素片辞書作成部211の構成をより詳細に示している。

#### 【0022】

図4において、401は音声素片検索部、402は音声素片保持部、403はHMM学習部、404はHMM保持部、405は素片認識部、406は認識結果保持部、407は登録素片決定部、408は登録素片保持部である。なお、210は図2で示した音声データベースである。

#### 【0023】

音声素片検索部401は、音声データベース210から所定の音韻環境を満たす音声素片を検索する。ここでは複数の音声素片が検索される。音声素片保持部402は、それら検索された音声素片を複数保持する。HMM学習部403は、音声素片保持部402に保持した各音声素片のケプストラムを、音声素片が有している波形の周波数情報をフーリエ変換等により求め、その結果に基づいて各音韻のHMMを求めて出力する。HMM保持部404は、その学習結果（HMMモデル）を各音韻ごとに保持する。素片認識部405は、学習したHMMを用いてHMMの学習に使用した全ての音声素片を素片認識し、尤度が最大となるHMM（最尤HMM）を求める。そして、現在注目している音声素片が最尤HMMの学習に使用した音声素片か否かを判断する。認識結果保持部406は、その素片認識結果を保持する。登録素片決定部407は、素片認識部405における認識結果から、素片認識に成功した音声素片のみを登録素片として採用する。登録素片保持部408は、登録素片決定部407により素片辞書206に登録する音声素



片のみを保持する。

【0024】

図5は、本実施の形態に係る素片辞書作成モジュールの動作を示すフローチャートである。

【0025】

ステップS501では、例えば、d i p h o n（ダイフォン）を音韻単位とする全ての音韻について処理を行なったかどうかを判定し、未処理の音韻が存在する場合はステップS502に進むが、未処理の音韻が存在しない場合はステップS504の素片認識処理に進む。

【0026】

ステップS502では、音声素片検索部401が、各音韻毎に、音声データベース210から所定の音韻環境を満足する音声素片を検索し、その検索された複数の音声素片を音声素片保持部402に保持してステップS503に進む。ステップS503では、HMM学習部405が、検索された複数の音声素片を学習データとして、ある音韻のHMMの学習を行なう。具体的には、まず音声波形のサンプリングレート22050Hzから、ケプストラム16次、デルタケプストラム16次、及びパワー、デルタパワーの計34次元のベクトルを、窓幅25.6m秒でフレーム幅2.5m秒ごとに求める。但し、ここではパワーとデルタパワーの値は、音声データベース210の文ごとに“0”から“1”に正規化されているものとする。次に、5状態1混合分布のHMM初期モデルを作成し、上記条件で得られるケプストラム・ベクトルを使ってHMMの学習を行なう。この学習の結果得られたある音韻のHMMを、HMM保持部404に保持した後、次の音韻のHMMを求めるため、ステップS501に戻る。

【0027】

ステップS504では、素片認識部405が、各音韻のHMMを用いて、ステップS502で検索された全ての音声素片に対して素片認識を行なう。つまり、各音声素片毎に、音声素片と各音韻のHMMとの尤度を求める。次にステップS505に進み、音声素片毎に、音声素片と最も尤度が高いHMMを求め、その音声素片が、そのHMMの学習に用いた音声素片であるかどうかを判定し、そうで

あれば素片認識が成功したとみなしてステップ S 5 0 6 に進み、その音声素片を素片辞書 2 0 6 に登録する。

【 0 0 2 8 】

一方、ステップ S 5 0 5 で、違うと判別された場合はステップ S 5 0 7 に進み、素片辞書 2 0 6 に登録しないと決定し、素片辞書 2 0 6 に登録することなくステップ S 5 0 8 に進む。こうしてステップ S 5 0 6 或いは S 5 0 7 の処理を実行した後ステップ S 5 0 8 に進み、ステップ S 5 0 4 で、全ての音韻の HMM の学習に用いた全ての音声素片について判定処理が終了したかを判定し、終了していなければステップ S 5 0 5 に戻って前述の処理を実行する。

【 0 0 2 9 】

以上説明したように本実施の形態 1 によれば、所定の音韻環境を満足する複数の音声素片を用いて各音韻に対応する HMM を学習し、学習した HMM を用いて HMM の学習に使用した全ての音声素片を素片認識し、最尤 HMM の学習に使用した音声素片であると判定された音声素片のみを素片辞書に登録する。

【 0 0 3 0 】

このように構成することにより、異音や雑音を含む音声素片を除去した素片辞書を生成することができ、合成音声の音質劣化を抑制することのできる素片辞書を提供できる。またこのような手順に寄って生成された素片辞書 2 0 6 を用いて音声を合成することにより、合成音声の音質の劣化を抑制することができる。

【 0 0 3 1 】

〔実施の形態 2〕

上述の実施の形態 1 においては、HMM 学習部 4 0 2 が、各音韻毎に HMM を作成し、素片認識部 4 0 5 では、各 HMM の学習に用いた全ての音声素片を対象に尤度を求める場合を説明したが、本発明はこれに限定されるものではない。例えば、音韻として d i p h o n e を採用する場合は、各音韻を C C 型、C V 型、V C 型、V V 型の 4 種類に分類して、同じ分類に属する音声素片を対象に素片認識を行なっても良い。ここで、C は子音、V は母音を表している。

【 0 0 3 2 】

〔実施の形態 3〕

また、前述の実施の形態 1, 2 において、認識に失敗した音声素片を登録しない場合で説明したが、本発明はこれに限定されるものではなく、例えば、許容できる認識誤りパターンを予め記述したテーブルを用意しておき、認識に失敗した音声素片であっても、そのテーブルに用意されている許容パターンに該当する場合は、登録素片決定部 4 0 7 において、素片辞書 2 0 6 に登録しても良いと判定する。

## 【 0 0 3 3 】

図 6 は、本実施の形態 3 に係る許容テーブルの一例を示す図である。

## 【 0 0 3 4 】

図 6 では、例えば、音韻として d i p h o n e を採用した例を示す。この場合、d i p h o n e 「a. y」の HMM の学習に用いた音声素片が「a. i」として認識された場合であっても、また、d i p h o n e 「a. k」の HMM の学習に用いた音声素片が「a. p」或いは「a. t」と認識された場合であっても、許容できるものとして素片辞書 2 0 6 に登録する。

## 【 0 0 3 5 】

この場合の処理を図 7 のフローチャートで示す。この処理は、図 5 のステップ S 5 0 5 において、違うと判別された場合に実行され、まずステップ S 6 0 1 に進み、許容テーブル（登録素片決定部 4 0 7 に設けられている）を検索して、その認識した結果が、そのテーブルに登録されているかどうかをみる。登録されているときは図 5 のステップ S 5 0 6 に進んで、素片辞書 2 0 6 に登録するが、一致しないときはステップ S 5 0 7 に進んで素片辞書 2 0 6 に登録しない。

## 【 0 0 3 6 】

## 〔実施の形態 4〕

上述の実施の形態 2 において、音韻として d i p h o n e を採用する場合において、認識に失敗した音声素片は登録しない場合について説明したが、本発明はこれに限定されるものではなく、認識に成功した素片数が閾値以下の音韻については、例えば V C 型の場合は、V 部分が一致していれば許容するようにしてもよい。

## 【 0 0 3 7 】

## 〔実施の形態 5〕

上述の実施の形態 1 において、各音声素片につき、ステップ S 5 0 3 で求めた全ての音韻の HMM との尤度をそれぞれ求める場合について説明したが、本発明はこれに限定されるものではない。ある音韻の HMM とその HMM の学習に用いた音声素片との尤度を求め、最も尤度の高い音声素片から上位 N 個（N は整数）を登録するか、或いは、所定の閾値以上の尤度を持つ音声素片だけを登録するようにしてもよい。

## 【 0 0 3 8 】

## 〔実施の形態 6〕

実施の形態 1 ～ 5 において、ステップ S 5 0 4 で求めた尤度を正規化せずに比較する場合について説明したが、本発明はこれに限定されるものではない。各尤度を、その尤度に対応する音声素片の時間長で正規化し、その正規化した尤度を用いて、登録する音声素片を、上述の手順により選択しても良い。

## 【 0 0 3 9 】

上記実施の形態においては、各部を同一の計算機上で構成する場合について説明したが、本発明はこれに限定されるものではなく、ネットワーク上に分散した計算機や処理装置などに分かれて各部を構成してもよい。

## 【 0 0 4 0 】

また上記実施の形態においては、プログラムを制御メモリ（ROM）に保持する場合について説明したが、これに限定されるものではなく、外部記憶など任意の記憶媒体を用いて実現してもよい。また、同様の動作をする回路で実現してもよい。

## 【 0 0 4 1 】

なお本発明は、複数の機器から構成されるシステムに適用しても、1 つの機器からなる装置に適用してもよい。前述した実施の形態の機能を実現するソフトウェアのプログラムコードを記録した記録媒体を、システム或いは装置に供給し、そのシステム或いは装置のコンピュータ（または CPU や MPU）が記録媒体に格納されたプログラムコードを読み出し実行することによっても達成される。

## 【 0 0 4 2 】

この場合、記録媒体から読み出されたプログラムコード自体が前述した実施の形態の機能を実現することになり、そのプログラムコードを記録した記録媒体は本発明を構成することになる。

【0043】

プログラムコードを供給するための記録媒体としては、例えば、フロッピーディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモ리카ード、ROMなどを用いることができる。

また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施の形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているOSなどが実際の処理の一部または全部を行ない、その処理によって前述した実施の形態の機能が実現される場合も含まれる。

【0044】

更に、記録媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行ない、その処理によって前述した実施の形態の機能が実現される場合も含まれる。

【0045】

以上説明したように本実施の形態によれば、HMMを用いた素片認識結果を利用して素片辞書に登録する音声素片を選択することによって、異音や雑音を含む音声素片を除去でき、音質の劣化が少ない合成音声を生成することを可能とする音声合成装置および方法を提供することができる。

【0046】

【発明の効果】

以上説明したように本発明によれば、HMMを用いた素片認識を行ない、その認識結果に応じて辞書に登録することにより、音声合成で再生される音声の品質を向上できるという効果がある。

【0047】

また本発明によれば、音声合成における音質の劣化を防止した素片辞書を作成できる。

【 0 0 4 8 】

またこのような素片辞書を用いて音声合成することにより、高品位に音声合成された音声を再生できる。

【図面の簡単な説明】

【図 1】

本発明の実施の形態に係る音声合成装置のハードウェア構成を示すブロック図である。

【図 2】

本発明の実施の形態 1 に係る音声合成装置のモジュール構成を示すブロック図である。

【図 3】

本実施の形態に係る音声合成モジュールにおける処理の流れを示すフローチャートである。

【図 4】

本実施の形態に係る素片辞書作成モジュールの詳細なモジュール構成を示すブロック図である。

【図 5】

本実施の形態 1 に係る素片辞書作成モジュールにおける処理の流れを示すフローチャートである。

【図 6】

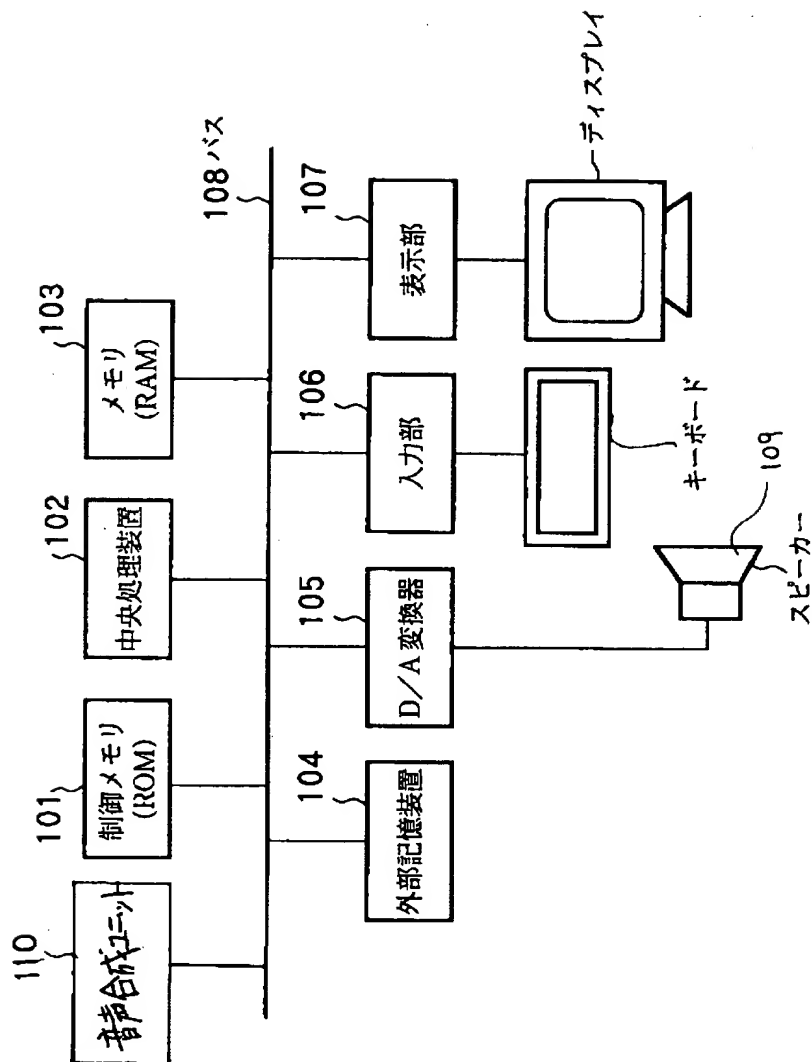
本発明の実施の形態 3 に係る誤認識許容パターンを記憶するテーブルの構成を示す図である。

【図 7】

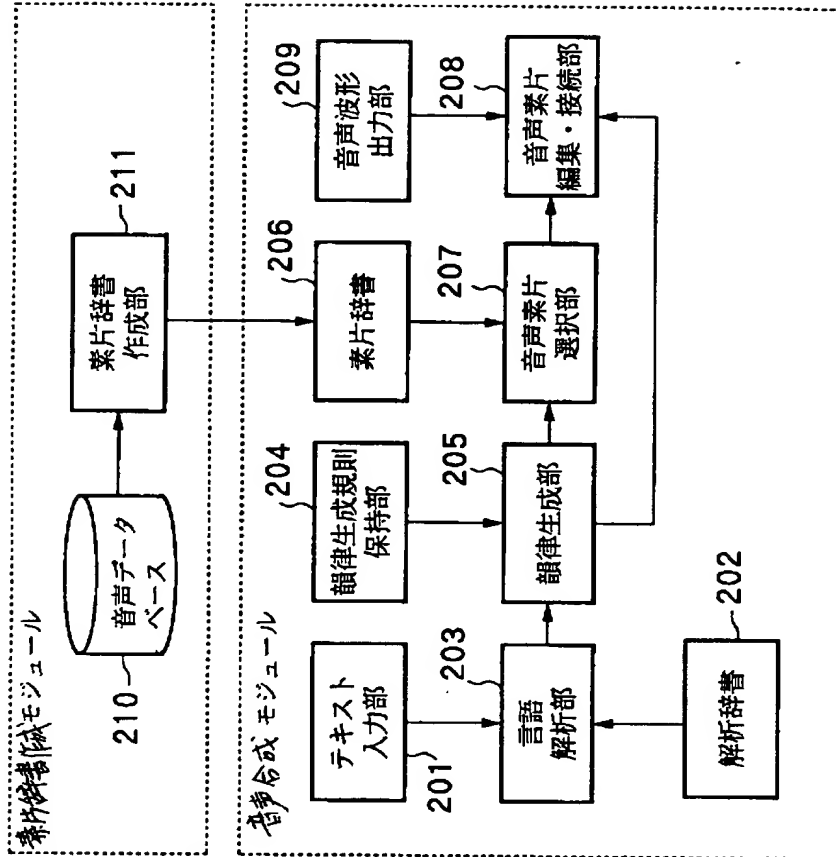
本発明の実施の形態 3 に係る素片辞書作成モジュールにおける処理の流れを示すフローチャートである。

【書類名】 図面

【図 1】

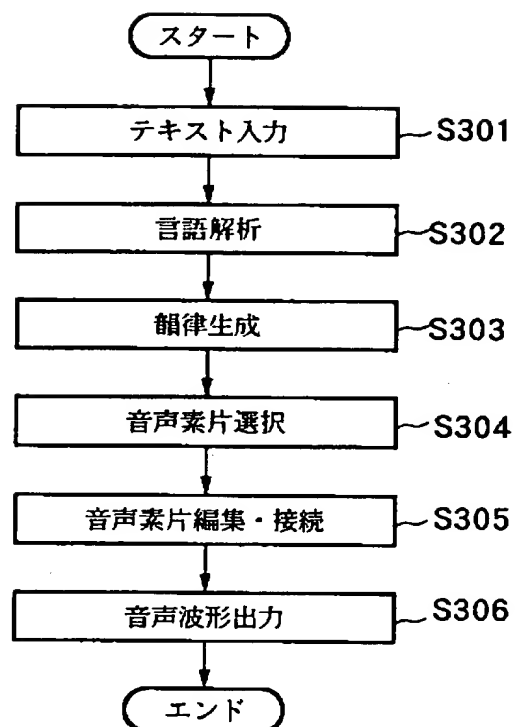


【図 2】

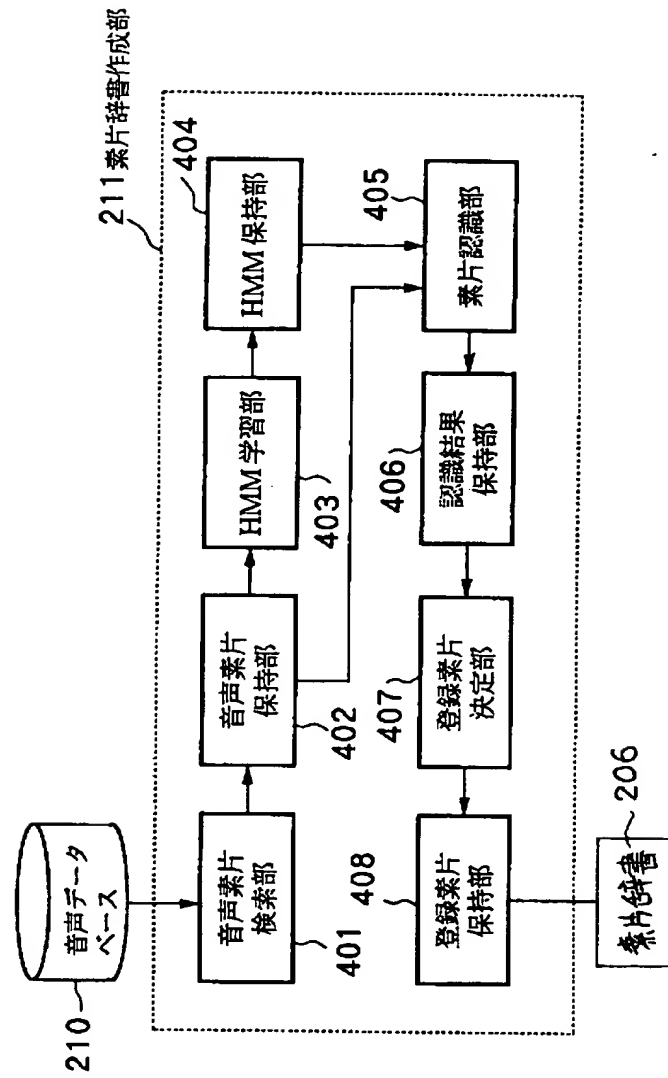




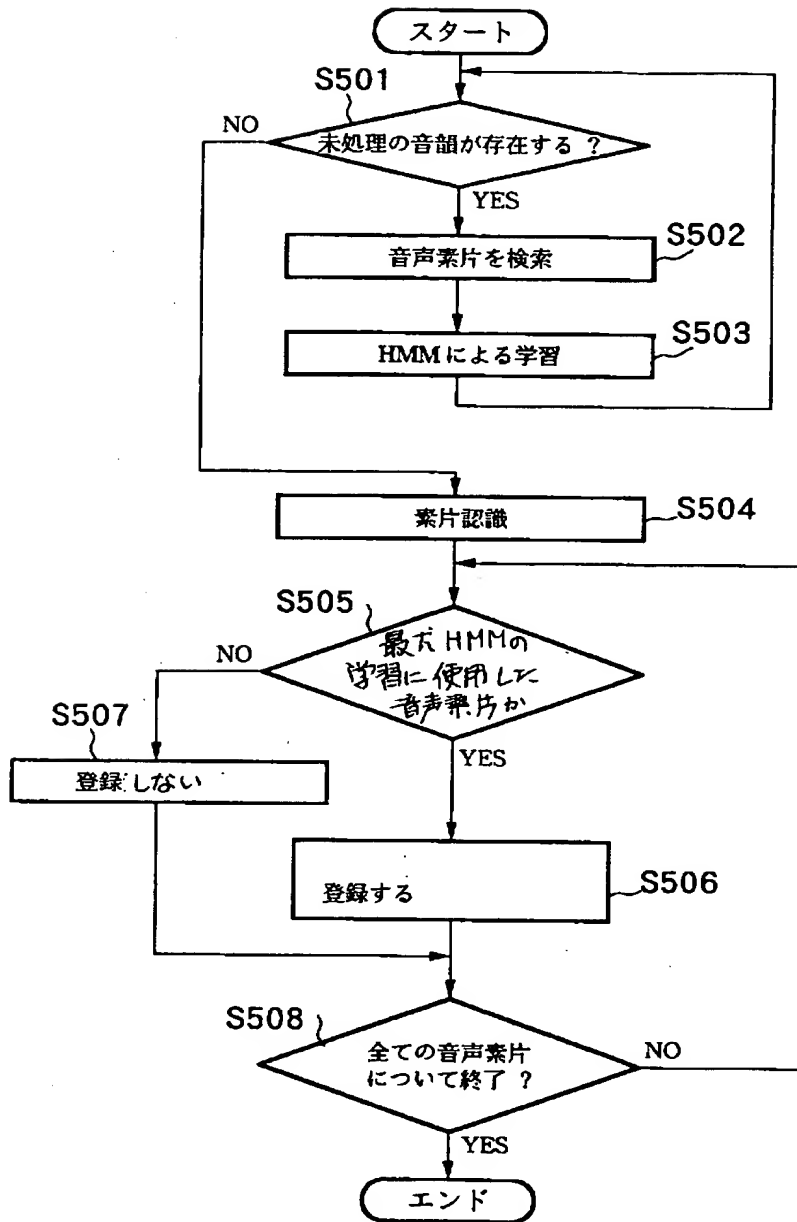
【図 3】



【図 4】



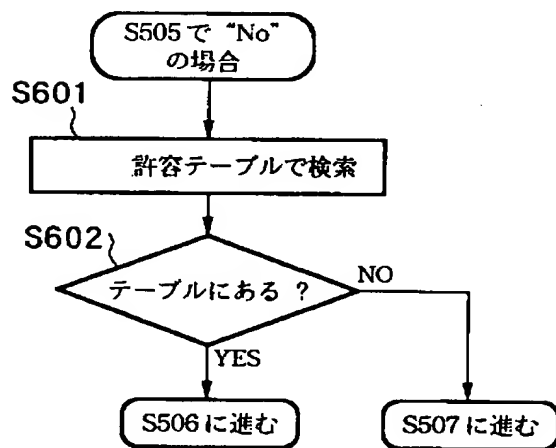
【図 5】



【図 6】

認識対象	許容パターン
a.y	a.i
a.k	a.p.a.t
...	
i.n	...
...	

【図 7】



【書類名】 要約書

【要約】

【課題】 音声合成における音質の劣化を防止した素片辞書を作成する。

【解決手段】 音声索引検索部 4 0 1 により、音韻環境を満足する音声素片を音声データベース 2 1 0 から検索し、その検索結果に基づいて HMM 学習部 4 0 3 において各音韻の HMM モデルを算出する。素片認識部 4 0 5 は、こうして算出された各音韻の HMM モデルに基づいて音声素片の素片認識を行ない、その素片認識結果が、元の音声素片と等しい場合に素片辞書に登録する。

【選択図】 図 4

出 願 人 履 歴 情 報

識別番号 [000001007]

1. 変更年月日 1990年 8月30日

[変更理由] 新規登録

住 所 東京都大田区下丸子3丁目30番2号  
氏 名 キヤノン株式会社